AMENDMENTS TO THE CLAIMS:

This listing of claims will replace all prior versions, and listings, of claims in the application:

1. (Currently amended) An apparatus for encoding a DNA sequence to achieve a high data compression ratio for storage or transfer, which comprises:

a comparative unit for aligning a reference sequence having known DNA information with a subject sequence to be ~~encoded~~ compressed and extracting a difference between the reference sequence and the subject sequence;

a conversion unit for converting the extracted difference between the reference sequence and the subject sequence into a string of characters and for outputting the string of characters,

wherein a type of the extracted difference comprises

a start region mismatch between the reference sequence and the subject sequence;

a blank representing there is no base in a base position in the subject sequence corresponding to the reference sequence;

a single base pair mismatch between the reference sequence and the subject sequence;

a base insertion into the subject sequence;

a multiple base pair mismatch between the reference sequence and the subject sequence, or

an end region mismatch between the reference sequence and the subject reference;

a code storage unit for storing a conversion code that corresponds to each character in ~~a~~ the string of characters to represent the extracted difference; and

an encoding unit for encoding the string of characters using the stored conversion codes.

2. (Currently amended) The apparatus of claim 1, wherein the characters to represent the extracted difference comprise

a character representing each DNA base,

a numeric character representing ~~a number of base positions that characterize a feature of the extracted difference~~a start position of the extracted difference, a length of the extracted difference, or a distance between the start position of the extracted difference and an end position of the extracted difference,

a character representing starting or ending of the extracted difference, and

a character representing ~~whether a type of extracted difference occurs in succession in the subject sequence~~continuation of the extracted difference.

3. (Currently amended) The apparatus of claim 2, wherein features of the extracted difference converted into the string of characters comprise

starting of the extracted difference,

a start position of the extracted difference,

~~whether a type~~continuation of the extracted difference ~~occurs in succession in~~ between the reference sequence and the subject sequence,

a number of continued bases in the extracted difference,

a base which the extracted difference comprises,

ending of the extracted difference, and

a distance between the start position of the extracted difference and the end position of the extracted difference.

4. (Canceled)

5. (Previously Presented) The apparatus of claim 1, wherein the conversion codes are 4 bit codes.

6. (Previously Presented) The apparatus of claim 1, which further comprises

a division unit for dividing the extracted difference into segments of a predetermined size, and

wherein the conversion unit converts the extracted difference into the string which is made up of the characters to represent the extracted difference based on the segments.

7. (Previously Presented) The apparatus of claim 1, which further comprises:

a compression unit for compressing the encoded subject sequence; and

a sequence storage unit for storing the compressed subject sequence.

8. (Previously Presented) The apparatus of claim 1, which further comprises

a pre-processing unit for modifying the reference sequence using a variation sequence generation factor created by a variation sequence generation function that uses random variables as inputs.

9. (Currently Amended) The apparatus of claim 8, wherein the variation sequence generation factor comprises a total number of variations, a distance between ~~two adjacent~~ variations, a length of a variation, a type of the variation, and a sequence of the variation.

10. (Withdrawn) A method for encoding a DNA sequence, which comprises:

aligning a reference sequence having known DNA information with a subject sequence to be encoded;

extracting a difference between the reference sequence and the subject sequence;

converting information of the extracted difference between the reference sequence and the subject sequence into a string of predetermined characters; and

encoding the individual characters that make the string of the predetermined characters using predetermined conversion codes that correspond to the individual characters.

11. (Withdrawn) The method of claim 10, wherein the characters comprises a first character representing DNA base symbols, a second character representing the number of the difference, a third character representing the starting and ending of the difference, and a fourth character representing continuation of the difference.

12. (Withdrawn) The method of claim 11, wherein converting comprises:

allotting the third character for the starting of the difference;

allotting the second character for the starting position of the difference;

allotting the fourth character for the continuation of the difference;

allotting the second character for the number of the continued bases of the difference;

allotting the first character for the bases of the difference;

allotting the third character for the ending of the difference;

allotting the second character for the distance between the start position and the end position of the difference; and

outputting the string of the allotted characters.

13. (Withdrawn) The method of claim 10, wherein the difference comprises start region mismatch between the reference sequence and the subject sequence, blank by base deletion of the subject sequence corresponding to the reference sequence, single base pair mismatch between the reference sequence and the subject sequence, base insertion into the subject sequence, multiple base pair mismatch between the reference sequence and the subject sequence, and end region mismatch between the reference sequence and the subject reference.

14. (Withdrawn) The method of claim 10, wherein the conversion codes are 4 bit codes, each of which corresponds to each of the characters.

15. (Withdrawn) The method of claim 10, which further comprises dividing the extracted difference into segments of predetermined sizes, and

wherein in converting, information of the extracted difference is converted into the string of the characters based on the segments.

16. (Withdrawn) The method of claim 10, which further comprises:

compressing the encoded subject sequence; and

storing the compressed subject sequence.

17. (Withdrawn) The method of claim 10, which further comprises, before aligning, creating a variation sequence induction factor from a variation sequence induction function that uses random variables as inputs and modifying the reference sequence using the created variation sequence induction factor.

18. (Withdrawn) The method of claim 17, wherein the variation sequence induction factor comprises the total number of variations, distance between the variations, length of the variations, type of the variations, and a variation sequence.

19. (Currently amended) A computer readable medium having embodied thereon a computer program for a method for encoding a DNA sequence to achieve a high data compression ratio, the method comprising:

aligning a reference sequence having known DNA information with a subject sequence to be encoded;

extracting a difference between the reference sequence and the subject sequence;

converting the extracted difference between the reference sequence and the subject sequence into a string of characters;

wherein a type of the extracted difference comprises

a start region mismatch between the reference sequence and the subject sequence;

a blank representing there is no base in a base position in the subject sequence corresponding to the reference sequence;

a single base pair mismatch between the reference sequence and the subject sequence;

a base insertion into the subject sequence;

a multiple base pair mismatch between the reference sequence and the subject sequence, or

an end region mismatch between the reference sequence and the subject reference; and

encoding each character in the string of characters using a conversion code ~~that corresponds to a character~~corresponding to the character,

wherein the computer readable medium ~~is not a carrier wave~~ comprises a ROM, a RAM, a CD-ROM, a magnetic tape, a floppy disk, or an optical storage medium.